REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
14-02-2016	Final Report	15-Jun-2012 - 14-Jun-2015

4. TITLE AND SUBTITLE	
Final Report: 4.3.1 Integrated Intelligence: Robot Instruction via Interactive Grounded Learning	5a. CONTRACT NUMBER W911NF-12-1-0197
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER 611102
6. AUTHORS Luke Zettlemoyer, Dieter Fox	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES	8. PERFORMING ORGANIZATION REPORT NUMBER
University of Washington 185 Stevens Way Box 352500 Seattle, WA 98195 -2500	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES) U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211	10. SPONSOR/MONITOR'S ACRONYM(S) ARO 11. SPONSOR/MONITOR'S REPORT NUMBER(S) 62161-NS.7

12. DISTRIBUTION AVAILIBILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT

People communicate using a variety of modalities, including speech, gestures, gaze, and whole body motion. To be effective, robots must be able to interpret the rich information contained in human communication. The goal of this project is to allow people to instruct robots and to teach them about objects and attributes using a combination of modalities. To achieve this goal, we are developing an interactive grounded learning system that can interpret rich human input such as speech, gesture, body motion, and gaze; and learn to identify objects based on the names and

15. SUBJECT TERMS

Robotics; Natural Language Processing; Grounded Language Learning; Learning Visual Attributes

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
UU	UU	UU	UU		Luke Zettlemoyer
					19b. TELEPHONE NUMBER 206-685-1227

Standard Form 298 (Rev 8/98)
Prescribed by ANSI Std. Z39.18

Report Title

Final Report: 4.3.1 Integrated Intelligence: Robot Instruction via Interactive Grounded Learning

ABSTRACT

People communicate using a variety of modalities, including speech, gestures, gaze, and whole body motion. To be effective, robots must be able to interpret the rich information contained in human communication. The goal of this project is to allow people to instruct robots and to teach them about objects and attributes using a combination of modalities. To achieve this goal, we are developing an interactive grounded learning system that can interpret rich human input such as speech, gesture, body motion, and gaze; and learn to identify objects based on the names and attributes people use to refer to them. This progress report summarizes the final findings and describes completed work towards a full, interactive grounded learning system.

Enter List of papers submitted or published that acknowledge ARO support from the start of the project to the date of this printing. List the papers, including journal references, in the following categories:

(a) Papers published in peer-reviewed journals (N/A for none)

<u>Received</u> <u>Paper</u>

TOTAL:

Number of Papers published in peer-reviewed journals:

(b) Papers published in non-peer-reviewed journals (N/A for none)

<u>Received</u> <u>Paper</u>

TOTAL:

Number of Papers published in non peer-reviewed journals:

(c) Presentations

Number of Presentations: 0.00

Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

<u>Received</u> <u>Paper</u>

TOTAL:

Number of Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received		Paper
02/09/2016	8.00	Maxwell Forbes, Rajesh P. N. Rao, Luke Zettlemoyer, Maya Cakmak. Robot Programming by Demonstration with Situated Spatial Language Understanding, International Conference on Robotics and Automation (ICRA). 26-MAY-15, . : ,
02/14/2016	9.00	Eunsol Choi , Tom Kwiatkowski, Luke Zettlemoyer. Scalable Semantic Parsing with Partial Ontologies, Association for Computational Lingustics (ACL). 01-JUL-15, . : ,
02/14/2016	10.00	Yuyin Sun, Adish Singla, Dieter Fox, Andreas Krause. Building Hierarchies of Concepts via Crowdsourcing, Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI). 01-JUL-15, . : ,
09/05/2013	2.00	Luke Zettlemoyer, Yoav Artzi. Weakly Supervised Learning of Semantic Parsers for Mapping Instructions to Actions, Transactions of the Association for Computational Linguistics (Presented at NAACL 2013). 10-JUN-13, . : ,
09/05/2013	1.00	Yuyin Sun, Liefeng Go, Dieter Fox. Attribute Based Object Identification, IEEE International Conference on Robotics and Automation (ICRA). 06-MAY-13, . : ,
09/22/2013	3.00	Nicholas FitzGerald, Yoav Artzi, Luke Zettlemoyer. Learning Distributions over Logical Forms for Referring Expression Generation, Emperical Methods in Natural Language Processing (EMNLP). 18-OCT-13, . : ,
09/22/2013	4.00	Tom Kwiatkowska, Eunsol Choi, Yoav Artzi, Luke Zettlemoyer. Scaling Semantic Parsers with On-the-fly Ontology Matching, Emperical Methods in Natural Langauge Processing (EMNLP). 18-OCT-13, . : ,
09/22/2014	5.00	Yuyin Sun, Liefeng Bo, Dieter Fox. Learning to Identify New Objects, 2014 IEEE International Conference on Robotics and Automation (ICRA). , . : ,
09/22/2014	6.00	Cynthia Matuszek, Liefeng Bo, Luke Zettlemoyer, Dieter Fox. Learning from Unscripted Deictic Gesture and Language for Human-Robot Interactions, Proceedings of the Conference on Artificial Intelligence . , . : ,

TOTAL: **9**

Number of Peer-Reviewed Conference Proceeding publications (other than abstracts):

(d) Manuscripts

<u>Received</u> <u>Paper</u>

TOTAL:

Number of Manuscripts:

Books

<u>Received</u> <u>Book</u>

TOTAL:

<u>Received</u> <u>Book Chapter</u>

TOTAL:

Patents Submitted

Patents Awarded

Awards

Graduate Students

NAME	PERCENT_SUPPORTED	Discipline
Yuyin Sun	0.50	
Eunsol Choi	0.50	
Maxwell Forbes	0.10	
FTE Equivalent:	**1.10**	
Total Number:	**3**	

Names of Post Doctorates

NAME	PERCENT_SUPPORTED
FTE Equivalent:	
Total Number:	

Names of Faculty Supported

NAME	PERCENT_SUPPORTED	National Academy Member
Dieter Fox	0.08	
Luke Zettlemoyer	0.08	
FTE Equivalent:	**0.16**	
Total Number:	**2**	

Names of Under Graduate students supported

NAME	PERCENT_SUPPORTED
FTE Equivalent:	
Total Number:	

Student Metrics

This section only applies to graduating undergraduates supported by this agreement in this reporting period

The number of undergraduates funded by this agreement who graduated during this period: 0.00

The number of undergraduates funded by this agreement who graduated during this period with a degree in science, mathematics, engineering, or technology fields:...... 0.00

The number of undergraduates funded by your agreement who graduated during this period and will continue to pursue a graduate or Ph.D. degree in science, mathematics, engineering, or technology fields:...... 0.00

Number of graduating undergraduates who achieved a 3.5 GPA to 4.0 (4.0 max scale):...... 0.00

Number of graduating undergraduates funded by a DoD funded Center of Excellence grant for Education, Research and Engineering:...... 0.00

The number of undergraduates funded by your agreement who graduated during this period and intend to work for the Department of Defense 0.00

The number of undergraduates funded by your agreement who graduated during this period and will receive scholarships or fellowships for further studies in science, mathematics, engineering or technology fields:...... 0.00

Names of Personnel receiving masters degrees

NAME
Total Number:

Names of personnel receiving PHDs

NAME	
Cynthia Matuszek	
Total Number:	**1**

Names of other research staff

NAME	PERCENT_SUPPORTED
FTE Equivalent:	
Total Number:	

Sub Contractors (DD882)

Inventions (DD882)

Scientific Progress

Our scientific progress has centered around three projects that focus on different aspects of our proposed work. We describe each below, focusing on our achievements in the final reporting period.

Building Hierarchies of Concepts via Crowdsourcing

Hierarchies of concepts and objects are useful across many real-world applications and scientific domains. Task-dependent hierarchies are expensive and time-consuming to construct. They are usually built in a centralized manner by a group of domain experts. This process makes it infeasible to create separate hierarchies for each specific domain. Attempts to build hierarchies using fully automatic methods have failed to capture the relationships between concepts as perceived by people. The resulting hierarchies perform poorly when deployed in real-world systems. With the recent popularity of crowdsourcing platforms, such as Amazon Mechanical Turk (AMT), non-expert workers (the crowd) can be employed at scale and low cost, to build hierarchies guided by their knowledge. Most existing methods as well as methods employing domain experts usually generate only a single hierarchy aiming to best explain the data or the semantic relationships among the concepts. This ignores the natural ambiguity and uncertainty that may exist in the semantic relationships.

In this work, we introduced a novel approach to learning hierarchies over concepts using crowdsourcing. Our approach incorporates simple questions that can be answered by non-experts without global knowledge of the concept domain. To deal with the inherent noise in crowdsourced information, people's uncertainty, and possible disagreement about hierarchical relationships, we develop a Bayesian framework for estimating posterior distributions over hierarchies. When new answers become available, these distributions are updated efficiently using a sampling based approximation for the intractably large set of possible hierarchies. The Bayesian treatment also allows us to actively generate queries that are the most informative given the current uncertainty. New concepts can be added to the hierarchy at time point, automatically triggering queries that enable the correct placement of these concepts.

We investigated several aspects of our framework and demonstrated that it is able to recover high-quality hierarchies for real world concepts using AMT. Importantly, by reasoning about uncertainty over hierarchies, our approach is able to unveil confusion of non-experts over concepts, such as whether tomato is a fruit or vegetable, or whether the wrist is part of a person's arm or hand. We believe that these abilities are extremely useful for applications where hierarchies should reflect the knowledge or expectations of regular users, rather than domain experts. This work was published in IJCAI 2015 and the paper is uploaded as part of this final report.

Robot Programming by Demonstration with Situated Spatial Language Understanding

General-purpose robots that can manipulate everyday object could take on many useful tasks in human environments. However, programming such robots to robustly function in every possible environment, with every possible user is extremely challenging. Instead, our research seeks to develop robots that can be programmed by their end-users to function in their particular environment.

Programming by Demonstration (PbD) is an intuitive method that enables users to program new capabilities on a robot simply by demonstrating the desired behavior. Kinesthetic teaching (i.e., physically moving the robot through desired states) allows users to directly demonstrate desired arm configurations and motions at a fine grained level and is often the most efficient interface for PbD. However, it requires physical effort from the user and it is not accessible to persons with certain motor impairments. Furthermore, it requires the user to be co-located with the robot. In this paper, we present a natural language interface for programming a robot to perform manipulation tasks purely through verbal commands. We exploited the fact that task-oriented motion is often relative to landmarks in the environment and that it can be described with spatial language (e.g., object reference expressions and prepositions).

We implemented a set of robust parametrized motion procedures on a PR2 mobile manipulator and we presented a situated language understanding model that takes a natural language utterance and infers the intended procedure and its parameter instantiation. We demonstrated that our method allows programming different manipulation tasks, even with ambiguous language and is on par with kinesthetic programming in terms of efficiency. This work was published in ICRA 2015 and the paper is uploaded as part of this final report.

Scalable Semantic Parsing with Partial Ontologies

Recently, significant progress has been made in learning semantic parsers for large knowledge bases (KBs) such as Freebase (FB) (Cai and Yates, 2013; Berant et al., 2013; Kwiatkowski et al., 2013; Reddy et al., 2014). Although these methods can build general purpose meaning representations, they are typically evaluated on question answering tasks and are designed to only parse questions that have complete ontological coverage, in the sense that there exists a logical form that can be executed against Freebase to get the correct answer. In this project, we instead consider the problem of learning semantic parsers for open domain text containing
concepts that may or may not be representable using the Freebase ontology. Sich an approach has wide application, including

especially building robust parsers for understanding robot directed utterance in human-robot interaction scenarios.

Even very large knowledge bases have two types of incompleteness that provide challenges for semantic parsing algorithms. They (1) have partial ontologies that cannot represent the meaning of many English phrases and (2) are typically missing many facts. They include subjective or otherwise unmodeled phrases such as “relaxed” and “quake-hit.” Freebase, despite being large-scale, contains a limited set of concepts that cannot represent the meaning of these phrases. They also refer to entities that may be missing key facts. For example, a recent study (West et al., 2014) showed that over 70% of people in FB have no birth place, and 99% have no ethnicity. In our work, we introduce a new semantic parsing approach that explicitly models ontological incompleteness and is robust to missing facts, with the goal of recovering as much of a sentence’s meaning as the ontology supports. We argue that this will enable the application of semantic parsers to a range of new tasks, such as information extraction (IE) and robotic dialog systems, where phrases rarely have full ontological support and new facts must be added to the KB.

To better model such text, we presented a new semantic parser that builds logical forms with concepts from a target ontology and open concepts that are introduced when there is no appropriate concept match in the target ontology. Figure 2 shows examples of the meanings that we extract. Only the first of these examples can be fully represented using Freebase, all other examples require explicit modeling of open concepts. To build these logical forms, we follow recent work for Combinatory Categorical Grammar (CCG) semantic parsing with Freebase (Kwiatkowski et al., 2013), extended to model when open concepts should be used. We develop a two-stage learning algorithm: we first compute broad coverage lexical statistics over all of the data, which are then incorporated as features in a full parsing model. The parsing model is tuned on a hand-labeled data set with gold analyses.

Experiments demonstrated the benefits of the new approach. It significantly outperforms strong baselines on both a referring expression resolution task, where much like in the QA setting we directly evaluate if we recover the correct logical form for each input noun phrase, and on entity attribute extraction, where individual facts are extracted from the groundable part of the logical form. We also see that modeling incompleteness significantly boosts precision; we are able to more effectively determine which words should not be mapped to KB concepts. When run on all of the Wikipedia category data, we estimate that the learned model would discover 12 million new facts that could be added to Freebase with 72% precision.

Technology Transfer